DOCUMENT RESUME

ED 101 537 FL 004 666

AUTHOR O'Malley, Michael H.
TITLE The Use of Prosodic Units in Syntactic Decoding.
PUB DATE 13 Apr 73
NOTE 14p.; Paper presented at the Annual Meeting of the Acoustical Society of America (85th, Boston, Massachusetts, April 1973)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Deep Structure; Grammar; Intonation; Language Rhythm; *Linguistic Theory; *Phonological Units; *Phonology; Phrase Structure; *Sentence Structure; Speech; Structural Analysis; Suprasegmentals; Surface Structure; *Syntax
IDENTIFIERS *Prosody (Linguistics)

ABSTRACT

This paper focuses on linguistic prosodic units related to boundaries between syntactic units. Specifically, rules for predicting the location of such boundaries, and factors affecting their location, are discussed. Examples are given on how prosodies can be used for syntactic analysis. Addressing the question of prosodic units and their distribution, two theories, both based on a hierarchy of units, are contrasted. A third theory, suggested as a possible basis for further refinement and testing is stated as follows: an NP that is not a single unstressed pronoun ends with a phrase- or clause-level boundary; NP's with embedded clauses have boundaries before the clause. A study conducted to test this rule is described, and several flaws in the rule are pointed out. Areas in the study of prosody that need further research are pinpointed. (AM)

SUR Note: 80
NIC #16094

Michael H. O'Malley
University of Michigan
Phonetics Laboratory
1079 Frieze Bldg.
Ann Arbor 48104

ED101537

THE USE OF PROSODIC UNITS IN SYNTACTIC DECODING*

Michael H. O'Malley

This paper is about prosodic information and how information might be used to guide the syntactic analysis of spoken utterances. As part of a session devoted to linguistic units, it might be appropriate to start by listing the prosodic units for a particular dialect of American English. Unfortunately, there is no universally agreed upon set of prosodic units. It is not even clear that prosodic information is organized into categorical units in the way that phonemes such as p, t, and k seem to be.

Disagreements over prosodic units does not imply that prosodies are unimportant. It is clear that the prosodic features of an utterance - its juncture, stress, intonation contour and rhythm - are determined in part by the grammatical structure of that utterance. In writing, punctuation, function words and inflectional morphemes are all used to signal syntactic structure. In speech, function words and morphemes tend to be unstressed and thus they are less intelligible. Greater emphasis must be placed on prosodies in order to signal the syntactic information. In fact there is evidence that prosodies are such good signals of syntactic structure, that on some dimensions, speech can be more syntactically complex than writing.

Pragmatics, which is that aspect of an utterance having to do with the speaker's attitude, interest and intention, also determines prosodic patterns. For example, prosodic features can act to focus attention on those portions of an utterance which the speaker thinks of as especially important to his message. The division of an utterance into phonological phrases and the placement of accents within those phrases are both functions of syntax and pragmatics.

Unfortunately some additional factors such as nervousness, thought processes and emotions can affect prosodic patterns, or rather, can produce effects, such as hesitation pauses, which are acoustically similar to prosodies. In general, these effects will be treated as noise to be overcome by a syntactic analysis system.

*Presented at the session on Linguistic Units at the 85th Meeting of the Acoustical Society of America in Boston on April 13, 1973.

FL 004 669

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

All in all, prosodic patterns are probably the most prominent features of spoken language. Children learn to understand and produce many of these patterns well before they develop a significant portion of their segmental phonology. Imitation of a language or imitation of a particular speaker is usually just imitation of the prosodic patterns which are characteristic of that language or of that individual. Even in a noisy environment, the prosodic features may be understood and used as an aid for decoding the spoken message. I am sure you have all had the experience of listening to a barely intelligible conversation, at first tuning into the prosodic patterns and only then being able to understand the words.

It should not be necessary to persuade people in the area of language that prosodies are important. But in spite of their importance, until quite recently only a very small fraction of the work which was reported at the Acoustical Society dealt with prosodies. There is a good reason for this neglect. Prosodies do not behave like other linguistic units such as phonemes and words.

The principle characteristic of such segmental units as words and phonemes is their discreteness. A word either is or is not a part of a particular utterance and a phoneme is or is not realized by a particular segment. Such discrete units can be studies by constructing paradigms of contrastive utterances. As the physical signal is varied across a unit boundary, the perception of one utterance versus another switches categorically. This categorical perception means that with only minimal training, speakers of a dialect can be taught to transcribe the words or phonemes in an utterance quite consistently.

Of course there are many exceptions to what I have said, but in general, arguments about segmental phonology are quite well understood. Such is not the case for prosodies. While some prosodic features of an utterance are quite clear, even experienced investigators will not agree in all aspects of a transcription. They will not agree with each other and they will not agree with an earlier transcription of their own. This disagreement is especially apparent in the transcription of spontaneous speech. If experienced investigators disagree about the units or even the dimensions of the units that they are studying, it is not surprising that progress in the field has been slowed.

This paper is about how prosodic information could be used to guide syntactic analysis. This question of how to use prosodic information, while it provides a framework for research, is really secondary. The fundamental

questions are: What are the prosodic units? Where do they occur? How reliably do they occur there? And can they be distinguished from each other and from other acoustic events?

If these could be answered, then the practical problem of actually using prosodies could be left to the cleverness of a computer programmer. The parsing strategy that is adopted would depend first upon the rules which predict the distribution or occurrence of prosodic elements, second upon the statistical reliability of these predictions for a particular type of speech and third, the strategy would depend upon the overall system organization.

The focus of this paper is on higher level, linguistic prosodic units - especially those which are related to boundaries between syntactic units. I am primarily concerned with rules for predicting the location of these boundaries and with factors which can modify their location.

Before discussing prosodic units and their distribution, I would like to present two examples of how prosodies might be used for analyzing syntax.

Suppose that you had found the following content words, as shown in line (1) on the handout.

(1) process computing average values used

Some of the possible readings for such a string are listed below it along with the prosodic breaks which would be likely to occur. In written English, the beginnings of units are generally marked by function words, but the ends are difficult to find. In speech, function words are often unstressed, but the ends of units seem to be more strongly marked by prosodies. The present example is quite clear. If lines (2) or (3) were the original utterance, we would all expect a rather well marked boundary after either computing or average - probably a silence interval as well as a fall-rise pitch contour and a lengthening of the preceding word. The problem is to make an algorithm which will detect a boundary where there is one and not have any false alarms elsewhere in the sentence.

Examples (6) and (7) illustrate one of the most common cases of ambiguity - a noun phrase followed by a prepositional phrase. The problem is whether the prepositional phrase modifies the noun it follows or some higher level structure. In writing, semantics must usually be invoked to resolve the ambiguity, but in speech, we might hope for a break between the noun phrase and the prepositional phrase when the prepositional phrase does not modify the noun phrase. The

presence of such a break might be predicted either by the branching of the surface structure tree or by the fact that the word block ends a noun phrase in case (7) but not in case (6). If breaks after noun phrases are sufficiently reliable, a left to right parser could use the information provided by the break to close off the noun phrase and start looking for a new structure. Even statistical tendencies could be used to change the order in which the parser explored alternatives.

While ambiguous sentences may be the most convincing argument for the necessity of prosodies in parsing, reliable prosodic signals, even in unambiguous sentences, such as (8) and (9), would be useful for guiding a syntactic parser.

Ilsa Lehiste has domonstrated that some speakers and listenerners can disambiguate some of the classical linguistic examples of ambiguous sentences. She has thus shown that the mechanism for indicating surface grouping is part of our linguistic knowledge. However, it is still necessary to investigate just how much of the ability to indicate surface grouping is actually used in spontaneous utterances.

Notice, of course, that surface structure grouping, as represented in tree diagrams, is hierarchical. There is no limit on the number of units which can be put inside other units. If higher level prosodic units are to reflect this grouping, then we might expect these units to be hierarchical also. However, there is not a separate prosodic unit for each node in the tree. There is not even a very close correspondence between the tree nodes and the prosodic units.

It is clear from these examples that boundaries or junctures play an important role in guiding a syntactic parse. The primary cue for this boundary seems to be a change in tempo or a slowing down which, for a sufficiently strong break, can become a pause or physical silence. Pause location has been studied a great deal, especially in the psycholinguistics literature. It is an easily measured parameter and does, in the case of slow and deliberate speech, provide information about the strength of the boundary. However, a pause which represents a grammatically determined boundary is easily confused with a hesitation pause which represents a nonlinguistic factor.

Hesitation pauses can be divided into silences, pauses filled with a sound such as 'uh' and false starts. The amount and type of hesitation depends on the individual's personality. Spontaneous speech, of course, has much more hesitation than read or practiced speech. In general, the syllable before the

hesitation is not lengthened, as it is for a grammatical pause, and the pitch contours are different. Hesitations tend to come early in a phrase and before less predictable words - rather as if the speaker is thinking of what to say next. They often cause the preceding word, which might be a function word such as the, to be stressed.

Hesitation pauses will undoubtedly pose a problem for any syntactic analysis system. Either they must be recognized and discarded or the system must be designed so that they will not hurt it. A system based only on silence duration will not be very robust in rejecting hesitation pauses.

I have given some examples of how prosodies might be used but I have so far avoided the question of prosodic units. Actually, there are several different theories, each with its own set of units, which could be described. I am going to outline a theory based on the British tradition of prosodic analysis and then contrast it with a simplified version of a theory due to Kenneth Pike. I am primarily concerned with units which span segments longer than a syllable; units whose function seems to be to indicate syntactic and semantic grouping.

Example (10) shows a sentence analyzed according to a theory proposed by M.A.K. Halliday. In this theory, there is a hierarchy of 4 units - phoneme, syllable, foot and tone group. At the highest level, an utterance is divided into a sequence of tone groups. A tone group corresponds very roughly to a clause. It usually contains, according to Halliday, a single unit of information. A non-restrictive relative clause, such as (11), would have two units of information or two assertions and thus two tone groups.

Tone groups are subdivided into feet. Each foot starts with a stressed syllable or with a silent 'beat'. Feet are perceptually isochronous in fluent speech. Feet are then further subdivided into syllables, all of which, except the first, are not stressed.

One foot in each tone group - called the tonic foot - is especially prominent. This foot represents, according to Halliday, the "information focus" in the tone group. The neutral or unmarked location for the tonic is on the last lexical item in the tone group. The tonic corresponds to what has been called "sentence" stress.

Each tone group has one or another intonation contour from the tonic to the end of the tone group. The primary pitch movement in the tonic is said to be on the tonic syllable.

The intonation breaks or junctures which, as we saw earlier, tend to occur at the ends of noun phrases do not necessarily correspond to tone group boundaries. If the break is strong enough to have an accompanying pause, then it can be marked as in sentence (12). However, I see no way to distinguish this transcription from that of a hesitation pause as in sentence (13). It seems to me that there needs to be a phrase-like unit in Halliday's system in between the foot and the tone group.

The phonological theory of Kenneth Pike is based on a hierarchy of units. In this theory, an utterance is simultaneously a sequence of units at several levels. An utterance is a sequence of phonemes, syllables, phonological words, phonological phrases, phonological clauses, phonological sentences, etc. The phonological word is a primary intonation contour which normally has a single stressed syllable. The number of phonological words in an utterance thus corresponds roughly to the number of feet as described in Halliday's system. However, the boundaries between phonological words do not usually occur before stressed syllables and thus do not correspond to foot boundaries. Another difference between feet and phonological words is that several stressed syllables may be "unitized" into a single phonological word. Phonological words do not, of course, correspond very closely to grammatical words.

Pike's phonological clauses seem to correspond roughly to Halliday's tone group. In the normal or "unmarked" case they may line up with the grammatical clauses in an utterance, but this is by no means required.

In between the phonological word and clause is the phonological phrase. Pike recognized that there seemed to be a rhythmic unit that was larger than the primary contour but smaller than the clause. It would seem that this intermediate sized unit could account for the break between the noun phrase and the prepositional phrase which was described in some of the preceding examples. An example of a transcription according to Pike is given in (15).

Such a hasty discussion of phonological systems does a great disservice to the theories. However, I would like to sum up my views of prosodic units as follows:

In fluent speech there is a unit which usually has one stress, which is normally less than one second long, which receives a rhythmic "beat" and which usually consists of one or more content words. Such a unit might be called a phonological word.

There is also a larger unit which often matches a full sentence or clause. The end of this unit represents a major break in the utterance and is often accompanied by a silence interval. The end is normally marked by a decrease in tempo and by one of a small number of intonation patterns. Semantically, this unit, which will be called a phonological clause, often contains one piece of information or one assertion.

Between these two units, there seems to be a third rhythmic unit called a phonological phrase.

All three units involve tempo and intonation patterns. It might be the case that they just represent three points along a single dimension. However, I have some reason to believe that at least 3 layers between the syllable and the full utterance are needed. I know of nothing which shows that more than three layers can be perceived.

My evidence for at least two layers above the word level comes in part from an experiment with read algebraic expressions (O'Malley, 1973). In that experiment I found that if I recognized two lengths of pause, I could recover the parentheses and thus the tree structure for the expressions with a fair reliability. As another example, I have found that the break between and NP and a PP is signaled over 75% of the time, usually with a phonological phrase boundary. Finally, I have conducted some informal experiments in which Kenneth Pike has listened to acoustically distorted versions of a number of sentences. His fastest and most stable judgments seem to involve the location of phonological phrase boundaries, boundaries which he recognized as much by change in rhythm as by pitch. He seemed to then recognize word and clause bounaries in relation to the phonological phrase boundaries.

If there are not more than three prosodic layers of phonological units, then any syntactic grouping which speakers wish to communicate must be coded into these three layers. The model thus predicts a limit on depth of embedding and a limit on disambiguation.

In order to use prosodic units to guide syntactic analysis, it is necessary to know where these units occur - their distribution. We have seen that the surface-structure syntactic tree is related to prosodic boundaires. Thus we might expect that major breaks in the tree would result in junctures. We have also seen that grammatical phrases sometimes correspond to phonological units. Thus we might predict that NP's, for example, would begin and end with a juncture. Sentence number (16) disproves both of these overly simple theories. A single

syllable, unstressed pronoun, even if it does represent a high-level grammatical constituent, is not a candidate for a prosodic boundary.

The rule could be made more elaborate as follows: an NP which is not a single unstressed pronoun ends with a phrase or clause level boundary. NP's with embedded clauses have boundaries before the clause. An example, along with the predicted pauses in given in (17). Notice that I am not predicting junctures at the beginning of NP's. Note also that the predictive power of this theory is weakened by my uncertainty in assigning surface trees and recognizing embedded clauses as in the examples in (18).

This rule, even if valid, would not solve all the problems about prosodies. However, it would aid in the resolution of the most common type of syntactic ambiguity. In order to test the rule, I gathered a small sample of spontaneous speech of a type which might be used with a speech recognition system. I was not interested in rapid or emotional speech or speech in which interpersonal effects predominate, but rather in speech from a person who is thinking or trying to solve a problem while he is talking. I therefore collected protocols in three different situations. In the first task, subjects asked the experimenter to move various objects around in a block world of the type made famous by Terry Winograd. In the second, subjects were given a complex object made of small plastic shapes such as triangles and circles; they then instructed the experimenter to make a similar object. In the third task, subjects asked the experimenter to connect various electronic components so as to form a circuit.

Protocols were taken from two different subjects in each task. Ten sentences were then selected from each subject and all six subjects were asked to read the 50 sentences from the other subjects. These protocols are not a fair sample of English since, for example, they are almost all imperatives. However, they seem to be a reasonable smaple of careful, spontaneous speech. They do contain a considerable amount of interpersonal and inter-task variability.

The read speech provides an interesting source for comparison. In particular, since read speech has many fewer hesitations, grammatical pauses can be defined as those pauses which occur in both the read speech and in the spontaneous speech.

Two transcribers then marked the perceptual pauses in the 360 sentences. A third listener marked all points in the 60 spontaneous utterances where he heard a phonological phrase or clause boundary, even if not accompanied by a "pause". Spectrograms were also made of the 60 spontaneous sentences.

The 60 sentences were then analyzed according to the predictions of the rule as in (17). All places where the NP rule predicted a boundary were marked as were all strings of words which could be NP's but for which the rule did not predict a boundary.

Results in the form of a contingency table are shown in (19). The predicted boundaries for the spontaneous sentences are compared to the perceived pause. Results from the listener who also included junctures which are signaled only by rhythm and pitch movement are shown in (20).

In general, about 1/2 of the NP's which should not be followed by junctures do in fact have pauses. However, almost all of the NP's which should have junctures do. This means that in parsing, if you think you are at the end of a NP but you don't find a juncture, you aren't. However, if you do find a juncture, you probably are at the end. Thus even in spontaneous speech, it seems to be possible to eliminate the majority of ambiguous structures.

The numbers in this experiment should not be taken too seriously. There is no current speech system which can use prosodic rules to aid recognition. The utility of such rules can only be tested by their effect on the performance of a complete system. Furthermore, until the junctures are detected acoustically and the trees assigned automatically, the numbers are partially subjective.

The NP rule as formulated is much too simple. There are several other factors which influence juncture location. For example, utterance rate and utterance length are both important factors. If the tempo of an utterance is increased, some of the phonological boundaries seem to disappear. Also, there is a tendency for phonological units to be a certain length. If there is no syntactic break in a long utterance, phonological boundaries will appear anyway. In addition, transformational processes beyond the surface tree may produce boundaries (21).

An attempt to account for such factors has been made by Manfred Bierwisch. His rules are too complex to give here but their results are illustrated in the examples. Bierwisch's rules deal only with the surface tree. They ignore non-branching nodes and all node labels such as NP. The rules apply cyclically from the bottom to the top of the tree.

Bierwisch's first rule erases boundaries between unstressed words so as to produce phonological-word like units as in line (22). His next rule again cycles up the tree, erasing boundaries between these phonological words. The resulting units might sometimes correspond to Pike's phrases.

The number of boundaries erased depends on a parameter of tempo. Line (23) shows the "tempo" at which Bierwisch predicts that various boundaries will be erased. We are currently working on testing and further developing Bierwisch's rules.

In summary, at least some rather simple rules show promise of being quite reliable, even in spontaneous speech. I feel confident that the rule which I gave can be improved and put on an acoustic basis.

There are several areas in which further work is needed. Hesitation pauses must be separated from grammatical pauses. Clause, phrase and word level boundaries must also be found and distinguished from each other. Certain stress patterns - sequences of certain types of feet - seem to affect rhythm and to introduce extraneous pauses. If true, this phenomenon needs to be described. The whole question of rhythm is central to defining the units and detecting boundaries. We need a way to measure rhythm and especially, changes in rhythm. Finally, we need to know more about how the overall rate of an utterance and the lengths of its constituents affect its phonology.

Of course, the rules must be refined. They also need to be tested on data from other languages, especially those with significantly different surface trees such as Tamil or Japanese.

Since prosodies are so much a part of how we actually organize messages, I think it is important to study spontaneous speech as well as rehearsed speech. Prosodies are closely tied to syntax and semantics, so that it is also essential, when studying their phonology, to be aware of the syntactic tree and of any ambiguities. In fact, I think prosodies ought to be studied in the context of an automatic syntactic analysis system.

For a long time it has been recognized that prosodies are related to syntax, but only now is it becoming important to use prosodies to aid in syntactic analysis. I hope that this new interest will result in a better overall balance between research on segmentals and on prosodies. Perhaps it will also serve to lower some barriers between research in phonetics and research in syntax and semantics.

BIBLIOGRAPHY

Bierwisch, M. (1966). Regeln fur die Intonation deutscher Satze, Studia Grammatica, 7, 99-201.

Halliday, M.A.K. (1967). Intonation and Grammar in British English, The Hague, Mouton (Janua Linguarum Series Practica 48).

Lehiste, I. (1973). Rhythmic Units and Syntactic Units in Production and Perception, paper presented at the 85th Meeting of the Acoustical Society of America, Boston, April 13, 1973.

O'Malley, M.H., D. Kloker, B. Dara-Abrams (in press). Recovering Parentheses from Spoken Algebraic Expressions, IEEE Transactions on Audio and Electroacoustics.

Pike, K. (1945). Intonation of American English, Ann Arbor: University of Michigan Press.

Robinson, J. (1973). Predicting Phonological Phrases, SUR Note: 73, Phonetics Laboratory, University of Michigan, Ann Arbor.

M. O'Malley
April 13, 1973
85th ASA Meeting

(1) process computing average values used

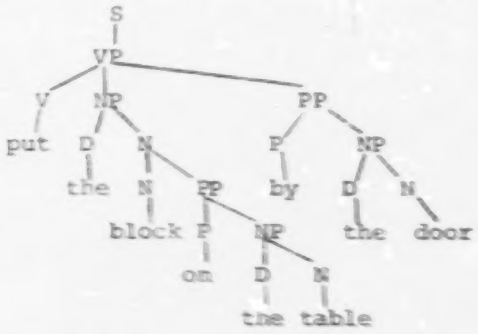
(2) In the process of computing | the average values will be used.

(3) In the process of computing the average | values will be used.

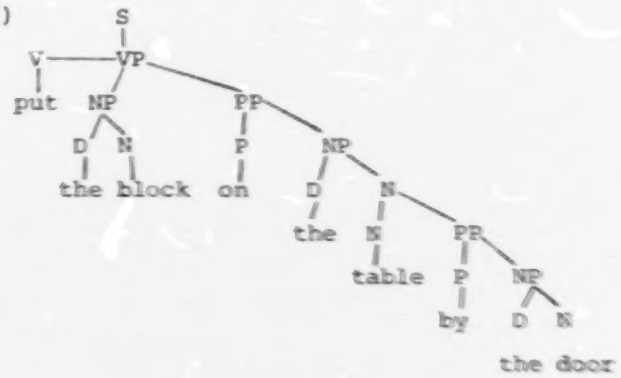
(4) The process of computing the average values | will be used.

(5) In the process | computing the average values will be used.

(6)



(7)



(8) Put the block near the door.

(9) In computing the average will be used.

(10) // can you / put$_{\wedge}$ / yellow / block on the / pyramid //
(foot; tone group; pause; tonic)

(11) // put it / near the / little / box // that you / just set / down //

(12) // put the / yellow / block$_{\wedge}$ in the / box on the / table //

(13) // put$_{\wedge}$ the / yellow / block$_{\wedge}$ in the / box on the / table //

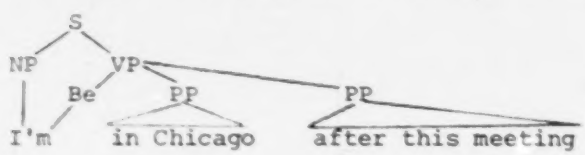
(14) Hierarchy of units:

Pike	Halliday
phoneme	phoneme
syllable	syllable
phonological word	foot
" phrase	tone group
" clause	
" sentence	
etc.	

(15) [(whát is the)(áverage)][(uránium)(léad rátió)]/[(for the lúnar sámples)]

( ) word [ ] phrase / clause ' stress

(16)



(17) Build a steeple with the blocks in the box that is by the door.
+ - +

(18) Did you find a flat space on top of the box.
?

A stack is two elements connected in parallel.
?

(19) Predicted \ Heard Pause

	Heard Pause −	Heard Pause +
pause −	36	27
pause +	28	89

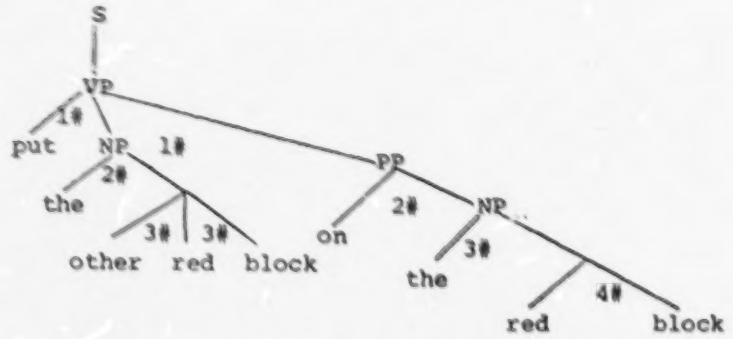
(20) Predicted \ Heard Juncture

	Heard Juncture −	Heard Juncture +
pause −	24	39
pause +	5	112

(21) John likes Mary and Mary, Dave.

---

put the ǒther réd blŏck | on the réd blŏck



0# put 1# the 2# ŏther 3# réd 3# blŏck 1# on 2# the 3# réd 4# blŏck 0#

(22) 0# put the ŏ.ther 3# réd 3# blŏck 1# on the réd 4# blŏck 0#

p=0 0# put the ŏ.ther réd blŏck 1# on the réd blŏck 0#
p=1 same
p=2 same
p=3 0# put the ŏ.ther réd blŏck on the réd blŏck 0#

(23) [put [the [other red block]]$_3$ [on [the [red block]]]]

(what is the average) (uranium) (lead) (ratio) (for the lunar) (samples)
2 0 0 3 0